DOCUMENT RESUME

ED 027 532 AL 001 768

By- Pacak, M.; Roberts, A. Hood

Outline of the Course in Automated Language Processing.

American Univ., Washington, D.C. Center for Technology and Administration.
Pub Date Apr 69
Note- 12p.
EDRS Price MF-$0.25 HC-$0.70
Descriptors- Bibliographies, College Curriculum, *Computational Linguistics, Computer Science, Course Content, *Course Descriptions, Course Objectives, Course Organization, *Curriculum Guides, Information Processing, Mathematical Linguistics, Reference Materials

The course in computational linguistics described in this paper was given at The American University during the spring semester of 1969. The purpose of the course was "to convey to students with no previous experience an appreciation of the growing art of computational linguistics which encompasses every use to which computers can be put in manipulation of natural language." Each of the 16 class sessions is briefly outlined and a number of articles and books for each class are listed for recommended reading. The majority of these references are available in published form or from the Clearinghouse for Federal Scientific and Technical Information in Springfield, Virginia. (JD)

THE AMERICAN UNIVERSITY

CENTER FOR TECHNOLOGY AND ADMINISTRATION

SYLLABUS 55.650

OUTLINE OF THE COURSE IN AUTOMATED LANGUAGE PROCESSING

M. Pacak
A. Hood Roberts

The purpose of the course is to convey to students with no previous experience an appreciation of the growing art of computational linguistics which encompasses every use to which computers can be put in manipulation of natural language.

The literature in the field of computational linguistics is not yet sufficiently stable to meet the needs of teachers and students. The rapid growth of interest in computational linguistics calls for training of many specialists in computer-oriented scientific as well as non-scientific fields in which at least elementary knowledge of computational linguistics is desirable. Perhaps in several years applied computational linguistics will be known to everyone who works with computers.

The importance of the automatic digital computer as a laboratory instrument and as a research tool is well known. It serves two major purposes: reduction of data and derivation of conclusions from theoretical premises. As far as computational linguistics is concerned, the intricate phenomena called natural language can be described only by theories of great complexity; computer simulation is being used to test the consistency of linguistic theories and to improve or modify the weak points of a theory.

The basic knowledge of computational linguistics is useful in any field in which the computer is called to interpret or analyze natural language such as indexing and classifying documents, extracting and abstracting pertinent documents or analyzing an author's style of writing in resolving issues of disputed authorship or in machine translation.

Another large question concerns language pathologies. They are tremendous in scope and difficulty, ranging from the problems of speakers with mental disorders or incipient mental disorders, or those resulting from some kind of trauma.

In general, computational linguistics postulates models by which the data of language can be examined and manipulated to determine the patterns underlying the variation and complexity of details that we find in all human activity, including language.

Some of the best papers have been published only in the form of semi-published reports. The collections of papers available as books are mostly conference proceedings which are too advanced for students who need an introduction to the field. The articles which appeared in four books, namely "Natural Language and the Computer", Ed. P. Garvin; "Readings in Automatic Language Processing", Ed. D. Hays; "Automated Language Processing", Ed. H. Borko; and "Computational Linguistics," D. Hays, were selected in order to meet the students' need.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

AL 001 768

They are generally comprehensible to the newcomer, and they cover the field as it exists today, taking a fairly broad view of computational linguistics. In addition, a set of articles published in different journals was selected as recommended reading. The majority of them are available at the "Clearinghouse for Federal Scientific and Technical Information".

Session 1

Definition and scope of computational linguistics in general

Machine translation, its history, reasons for it and practical results achieved.

Differences between word-for-word translation and syntactic translation

Machine-aided translation and its comparison with MT.

Natural language and some crucial problems of morphological, syntactic and semantic analysis.

Sessions 2 & 3

Natural language and artificial languages (COBOL; SNOBOL; COMIT); differences in parameters and scope.

Standards for natural language texts in machine-readable form.

Systems for computer-oriented transliteration of characters which are not available on computers (cyrillic and greek characters, mathematical symbols, formulas etc.).

Types of concordances

a) Monolingual
b) Bilingual
c) Full
d) Selective
e) Restrictive

The preparation of computer concordances and their use in computational linguistics

Concordances and transliteration systems

Session 4 -
Computational Lexicography

1. Comparison of traditional lexicography and computational lexicography

2. Physical characteristics of computer oriented dictionaries (cards, tapes, drums, photoscopic discs).

3. Methods for compilation of dictionaries (concordances, commercial dictionaries, reverse dictionaries).

4. Microglossaries and their usage in information retrieval.

5. Formats of computer-oriented dictionaries:

   a) Full form dictionaries
   b) Stem dictionaries
   c) Root dictionaries
   d) Phrasal dictionaries

6. Coding of dictionaries:

   a) Grammar codes and their function
   b) Semantic codes
   c) Contextual codes

7. Bilingual MT dictionaries

8. Description and comparison of computer oriented dictionaries produced by Harvard University; Georgetown; University of Texas; University of California; the Rand Corporation; The Bunker-Ramo Corporation; IBM; ITEK Corporation.

Session 5
Computational Morphology

1. Segmentation of word forms into their basic components (roots, stems, derivational morphemes, inflectional morphemes, prefixes)

2. Procedure for segmentation of word forms and establishment of classes of stem morphemes according to the principle of complementary distribution

3. Formal representation of morphological analysis in the form of trees, tables, matrices, logical formulas, flow-charts

4. The nature of affixing in written English; structural definition of affixes from multisyllable words; part-of-speech implications of affixes

5. Morphological transformations and their usefulness for information retrieval

6. Homographs: Definition and classification

7. Types of homography:

   a) Morphological
   b) Morphosyntactic
   c) Syntactic
   d) Semantic

8. Ambiguous sentences and their treatment in computational linguistics

9. Procedures for the resolution of homography

Session 6
Automated Speech Analysis and Synthesis

1. Phonological Analysis

   a) The problem of segmentation
   b) Phonemic versus graphemic codes

2. Speech Synthesis

   a) By formant
   b) By rule

3. Rate Control Recording

   a) Speech compression
   b) Speech expansion

Session 7
Quantitative Linguistics

1. Zipf's law

2. Enthropy and redundancy

3. Markov chains

Session 8
The Function of Grammars in Computational Linguistics

1. Notional grammars

2. Formal grammars, their properties and relation to machines

3. Major types of formal grammars:

   a) Unrestricted rewrite (Turing machine)
   b) Finite state grammars
   c) Categorial
   d) Phrase-structure
   e) Transformational

   < context-sensitive
   < context-free

4. Evaluation of grammars in terms of their usefulness for language data processing.

Session 9
Immediate Constituent Grammar

1. Phrase-structure grammars and Cock's parsing logic

2. "Parse" system for the automatic syntactic analysis of English developed by the Rand Corporation

3. Major components of "Parse"

   a) Glossary of coded English word forms
   b) Table of grammatical rules
   c) Resultant codes
   d) Programming system

4. YNGVE's "depth hypothesis" for features of English syntax

Session 10
Dependency and Other Types of Grammar

1. Dependency grammar: definition

2. Major components of dependency grammar:

   a) Terminal alphabet
   b) Non-terminal alphabet
   c) Assignment function
   d) Dependency rules and their form

3. Flow-chart for dependency parsing

4. Concept of multiple dependency

5. Dependency and phrase structure grammar; comparison

6. Fulcrum grammar and its basic properties

7. Components of the fulcrum grammar:

   a) Dictionary

8. Syntax parses
   - Preliminary
   - Minor syntax
   - Major syntax
   - Terminal

9. Concept of heuristic syntax

10. Description of flow-charting

11. Predictive grammar and its properties

12. Components of the predictive grammar:

    a) Dictionary of word classes
    b) Grammar table of predictions
    c) Form of prediction rules
    d) Push down store

13. Flow-chart

Session 11
Transformational-Generative Grammar

Transformational Grammar was originally introduced by Harris, Z. and later on elaborated by Chomsky, N.

Concept of generative grammar (formal deductive system whose terminal expressions are sentences in a given language)

Components of generative grammar:

a) Syntactic
b) Phonological
c) Semantic

Concept of deep structure (determination of the semantic interpretation of a sentence) and surface structure (phonetic interpretation)

Reasons for the development of transformational grammar (lack of phrase structure rules to generate complex sentences)

Types of transformations < obligatory / optional

Result of transformations are kernel sentences

Basic kernel - sentence types in English (NP + aux + Vi; NP + Aux + Vt + Np; NP + aux + be NP etc.)

Types of transformational rules:

a) Deletions (AB → A)
b) Adjunctions (A → A-B)
c) Permutations (AB → BA)
d) Substitutions (AB → CD)
e) Combinations of b and c rules

Notion of syntactic features and introduction of complex symbols representing a set of specified syntactic features.

Types of basic rules:

a) Branching rules:

b) Subcategorization rules — strict subcategorization; selective rules

c) Programming aspects of transformations

d) Evaluation of transformational grammar

Session 12
Implementation of Automatic Language Processing to Information Retrieval Systems

1. Automatic Indexing

2. Automatic Abstracting

Session 13 & 14
General Problems in Semantics

Session 15
Review

Session 16
Final Examination

RECOMMENDED READINGS

Session 1

Dostert, L. E.: Machine Translation and Language Data Processing: Computers and Automation; Vol. XII; No. 5, 1963

Howerton, P. W.: The Parameters in Machine Translation; Machine Translation; Vol. 6, 1961.

Lehmann, W. P.: An Experiment in Machine Translation; The Graduate Journal; University of Texas; Vol. VII, No. 1, 1965

Mel'Chuk, I. A.: Machine Translation and Linguistics; in Exact Methods in Linguistic Research; University of California Press; Berkeley, 1963

Krollman, F.: A Procedure to Provide an Automatic Translation AID; NSF 1965

General Report on Machine Translation; Georgetown University, 1963

Dostert, L. E.: Machine Translation and Automatic Language Processing; Vistas in Information Handling; Spartan Books; Washington, 1963

ALPAC Report

Schultz, L.: Language and the Computer; in "Automated Language Processing" ed. by H. Borko; John Wiley and Sons, Inc., New York 1967

Lehmann, W. P.: Computational Linguistics; in "Seminar on Computational Linguistics" ed. by A. Pratt, H. Roberts and K. Lewis; HEW; Public Health Service Publication No. 1716, 1967

Pendergraft, E. D.: "Translating Languages"; Automated Language Processing; ed. H. Borko.

Sessions 2 and 3

Hays, D.: Concordances; in Computational Linguistics; American Elsevier Publishing Co., New York, 1967

Johnson, D. C.: The Basic Punching Code for Russian Text; University of California, Clearinghouse, PB 168 884 HC

Luhn, H. P.: Keyword-in-Context Index for Technical Literature; Readings in Automatic Language Processing; ed. by D. G. Hays;

Ziehe, T. W.: Natural Language in Computer Form; Readings in Automatic Language Processing; American Elsevier Publishing Co., New York, 1966

Lamb, S., Gould, L.: Concordance from Computers; Yale University; Clearinghouse PB 168 434 HC

General Report on Machine Translation; Georgetown University, 1963

Session 4

Josselson, H. H.: Automatization of Lexicography; Proceedings of International Conference on Computational Linguistics, New York, 1965

Josselson H. H.: Research in Automatic Russian-English Scientific and Technical Lexicography; Wayne State University; Clearinghouse, PB 173 176 HC

Hays, D.: Dictionary Lookup; in "Computational Linguistics", American Elsevier Publishing Co., New York, 1967

Lamb, S. M., Jacobsen W. H.: A High-Speed Large Capacity Dictionary; Readings in Automatic Language Processing

Harper, K. E.: Dictionary Problems in Machine Translation; Natural Language and the Computer; Ed. P. Garvin, McGraw-Hill Book Co., New York, 1963

General Report on Machine Translation; Georgetown University

Session 5

Earl L. L.: Automatic Determination of Parts of Speech of English Words; Mechanical Translation, Vol. 10, No. 3, 4; 1967

Pratt, A., Pacak, M.: Identification and Transformation of Terminal Morphemes in Medical English; Methods of Information in Medicine; Heidelberg, Germany, August 1969

Pacak, M.: Computational Morphology; TA Informations; No. 1, 1966, Editions, Klinsieck, Paris, France

Pacak, M. Morphology in Terms of Machine Translation; Advances in Documentation and Library Science, Vol. III, Part 2; Interscience Publishers, Inc., New York 1961

Dolby, J. L.: The Application of English Word Morphology to Automatic Indexing and Extracting; Clearinghouse, AD 615 424

Earl, L. L.: Structural Definition of Affixes from Multisyllable Words; Mechanical Translation and Computational Linguistics; Vol. 9, No. 2, 1966

Pacak, M.: Homographs: Their Classification and Identification; Vistas in Information Handling, E. P. Howerton; Spartan Books, Washington, 1962

Session 6

Pacak, M.: Homographs: Their Classification and Identification; Vistas in Information Handling, Ed. P. Howerton, Spartan Books, Washington, 1962

Pacak, M., Henisz, B.: Homographs: Their Classification and Identification; ERIC Document Reproduction Service, Bethesda, Maryland - ED 024 916

Flanagan, J. L.: Speech Analysis, Synthesis and Perception; Berlin, 1965

Pike, K. L.: Phonetics; Ann Arbor, Michigan

Fant, G.: Speech Analysis and Synthesis; Royal Institute of Technology, Speech Transmission Laboratory, Report No. 26, 1962

Fry, D. B.: Automatic Speech Recognition; Proceedings of the 4th International Congress in Phon. Sci., Helsinki 1961, 's-Gravenhage, 384-399; 1962

Joos, M.: Acoustic Phonetics; Language, 24, suppl. 2, 1964

Session 7

Pierce, J. R.: Symbols, Signals and Noise: The Nature and Process of Communication; Harper & Row, New York, 1961

Kucera, H., Monroe, G.K: A Comparative Quantitative Phonology of Russian, Czech and German; Number Four in the Series; Mathematical Linguistics and Automatic Language Processing; American Elsevier Publishing Co., New York, 1968

Jones, P.E., Giuliano, V.E., Curtice, R.M.: Automatic Language Processing; American Data Processing Inc., Michigan, 1969

Zipf, G.K.: The Psycho-Biology of Language: An Introduction to Dynamic Philology; The MIT Press, Mass., 1965

Roberts, A. H.: A Statistical Linguistic Analysis of American English; Mouton & Co., 1965

Session 8

Bobrow, D.G.: Syntactic Theory in Computer Implementation; in Automated Language Processing; Ed. H. Borko; John Wiley and Sons, N. Y., 1967

Garvin, P.: Syntax in Machine Translation; Natural Language and the Computer; Ed. P. Garvin, McGraw-Hill Book Co, New York, 1963

Pacak, M.: The Function of Grammars in Computational Linguistics; National Institutes of Health Seminar Paper; June 1968

Session 9

Robinson, J.: Automatic Parsing and Fact Retrieval: A Comment on Grammar, Paraphrase and Meaning; Clearinghouse AD 432 036 HC

Robinson, J.: A System for Automatic Syntactic Analysis of English Text; Clearinghouse AD 621 310 HC

Robinson, J., Marks, S.: Parse: A System for Automatic Analysis of English Text; Clearinghouse AD 621 311 HC

Robinson, J.: Preliminary Codes and Rules for the Automatic Parsing of English; Clearinghouse AD 295 651 HC

Yngve, V.: The Depth Hypothesis; Proceeding of Symposia in Applied Mathematics, Vol. 12, 1960, Providence, American Mathematical Society, 1961

Gleason, H.A.,Jr.: Linguistics and English Grammar; Holt, Rinehart & Winston, Inc., 1965

Bach, E.: An Introduction to Transformational Grammars; Holt, Rinehart & Winston, Inc., 1964

Session 10

Rhodes, I.: A New Approach to the Mechanical Analysis of Russian; Mechanical Translation, Vol. 6, 1961

Kuno, S.: The Predictive Analyzer; Readings in Automatic Language Processing

Oettinger, A.G.: Mathematical Linguistics and Automatic Translation; Clearinghouse, PB 252 HC

Garvin, P.L.: The Place of Heuristics in the Fulcrum Approach to Machine Translation; Lingua, Vol. 21, North-Holland Publishing Co., Holland, 1968

Kuno, S.: Automatic Syntactic Analysis; in Seminar on Computational Linguistics

Session 11

Bach, E.: An Introduction to Transformational Grammars; McGraw-Hill, Co., New York, 1964

Chomsky, N.: Aspects of the Theory of Syntax; The MIT Press, 1965

Harris, Z. S.: String Analysis of Sentence Structure; Mouton Co., The Hague, 1962

Sager, N.: Report on the String Analysis Programs; Clearinghouse PB 170 585 HC

Joshi, A. K.: Transformational Analysis by Computer; in Seminar on Computational Linguistics

Session 12

Borko, H.: Indexing and Classification; Automated Language Processing

Wyllys, R.E.: Extracting and Abstracting by Computer; Automated Language Processing

Sharp, J.R.: Auto-indexing and auto-abstracting; in "Some Fundamentals of Information Retrieval"; London House and Maxwell; New York, 1965

Salton, G.: Automatic Phrase Matching; Readings in Automatic Language Processing

Session 13

Katz, J. J. and Fodor, J. A.: The Structure of a Semantic Theory; Language, Part I, Vol. 39, No. 2, 1963

Rubenstein, H.: Directions in Semantic Research; in Seminar on Computational Linguistics

Ceccato, S.: Mechanical Translation; The Correlation Solution; Defense Documentation Center for Scientific and Technical Information; AD 409 607

April, 1969